Weighted Least Squares Estimators
on the Frequency Domain
for the Parameters of a Time Series

by

Shean-Tsong Chiu[1]

Technical Report 86-22, September 1986.

[1]Mathematical Sciences Department, Rice University, Houston, Texas 77251.

			Form Approved OMB No. 0704-0188

Report Documentation Page

1. REPORT DATE **SEP 1986**	2. REPORT TYPE	3. DATES COVERED **00-00-1986 to 00-00-1986**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Weighted Least Squares Estimators on the Frequency Domain for the Parameters of a Time Series	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **21**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Weighted Least Squares Estimators on the Frequency Domain

for the Parameters of a Time Series

Abstract

A procedure for estimating the parameters of a time series is proposed. The estimate minimizes a criterion function which is the weighted sum of squares of the distances between the periodograms and the spectrum of the series. Under mild conditions, the estimate is shown to be strongly consistent. The asymptotic distribution of the estimate is also obtained. With a proper choice of weighting function, the estimate has the same asymptotic distribution as the one for the maximum likelihood estimate. Simulations were carried out to evaluate the performance of the estimate.

1. Introduction

We consider a stationary time series $X(t)$, $t=0,1,\ldots,T-1$, with mean zero and autocovariance function $c(u) = E[X(t)X(t+u)]$. When the autocovariance function satisfies

$$\sum_{u=-\infty}^{\infty} |c(u)| < \infty \tag{1.1}$$

the spectrum density function of the series $X(t)$ is defined by

$$f(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} c(u)\exp(-i\lambda u) . \tag{1.2}$$

This relation can be inverted to obtain the representation

$$c(u) = \int_{-\pi}^{\pi} \exp(iu\lambda)f(\lambda)d\lambda . \tag{1.3}$$

In this paper, we are interested in the situations in which the spectrum of the series depends on some unknown parameters. Here are some examples:

$$f(\lambda) = \frac{\Omega_0}{[1+(\lambda/\lambda_0)]^2} \tag{1.4}$$

$$f(\lambda) = \frac{\Omega_0}{[1+(\lambda/\lambda_0)^2]^{3/2}} \tag{1.5}$$

$$f(\lambda) = \beta_2^2 \exp(-\lambda^2 \beta_1^{-2}/2) \tag{1.6}$$

$$f(\lambda) = \beta_2 \exp(-|\lambda|/\beta_1) \tag{1.7}$$

$$f(\lambda) = \frac{\beta_2}{\lambda^5} \exp(-\beta_1/\lambda^4) \tag{1.8}$$

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1+b_1 e^{i\lambda} + \cdots + b_q e^{iq\lambda}|^2}{|1+a_1 e^{i\lambda} + \cdots + a_p e^{ip\lambda}|^2} \tag{1.9}$$

Models (1.4) and (1.5) were used in Aki (1967) to model the spectra of far field body-wave displacements of earthquakes. The parameter λ_0 is the corner frequency and Ω_0 is proportional to seismic moment. Seismologists are very interested in estimating the parameters λ_0 and Ω_0, which are important in the study of the source properties of earthquakes (cf Aki and Richards (1980)). Slutsky (1937) used model (1.6) in studies of economic time series. Model (1.7) has been used by Lumley and Panofsky (1964) for atmospheric turbulence, and by Whittle (1962) for agricultural spatial series. Pierson and Moskowitz (1964) proposed (1.8) as the spectrum of ocean waves caused by winds. The last example (1.9) is the spectrum of an autoregressive moving average process of order (p,q) (ARMA(p,q)). In the following discussion, we let $f(\lambda, \theta_0)$ represent the spectrum of the series $X(t)$, and we are interested in estimating the parameter θ_0.

The criterion function used in this paper is based on periodograms. The periodogram at frequency λ of a series $X(t)$ is defined as

$$I(\lambda) = \frac{1}{2\pi T} d_X^T(\lambda) d_X^T(-\lambda) \tag{1.10}$$

where $d_X^T(\lambda)$ is the finite Fourier transform of the series $X(t)$,

$$d_X^T(\lambda) = \sum_{t=0}^{T-1} X(t) \exp(-i\lambda t). \tag{1.11}$$

The motivation of the frequency domain approaches is from the observation that, under some

mild conditions, the periodograms of a series $X(t)$ on the Fourier frequencies $0 < \lambda_i = 2\pi i/T < \pi$ are asymptotically independently distributed according to exponential distribution with mean $f(\lambda_i)$ (cf Brillinger (1975)). Therefore, an "approximate maximum likelihood estimator" can be obtained by finding the θ that maximizes the function

$$L(\theta) = -\sum_\lambda \log(f(\lambda,\theta)) - \sum_\lambda I(\lambda)/f(\lambda,\theta) \qquad (1.12)$$

where the summation is over the Fourier frequencies in $(-\pi,\pi)-\{0\}$. Unless indicated otherwise, we use this convention throughout this paper. When the spectrum can be written as

$$f(\lambda) = (2\pi)^{-1}g(\lambda,\theta_0)\sigma_0^2, \qquad (1.13)$$

$g(\lambda,\theta)$ does not depend on σ^2. We could obtain an estimate of θ_0 by minimizing

$$L'(\theta) = \sum_\lambda I(\lambda)/g(\lambda,\theta) \qquad (1.14)$$

and estimate σ_0^2 by $\hat{\sigma}^2 = L'(\hat{\theta})/(2\pi T)$. The estimates which maximize $L(\theta)$ or $L'(\theta)$ were studied in Bloomfield (1973), Hannon (1973), Whittle (1961) Dzhaparidze (1974), Davies (1973), Robinson (1978) and Ibragimov (1967). Under various conditions, the estimates were shown to be asymptotically consistent and efficient. A similar estimation procedure was proposed by Taniguchi (1981), who considered the estimates which minimize the distance between the spectrum and the smoothed periodograms.

$I(\lambda)/g(\lambda,\theta)$ can be viewed as the periodograms of the series past a filter whose transfer function $A(\lambda)$ satisfies $|A(\lambda)|^2 = 1/g(\lambda,\theta)$. The problem of minimizing $L'(\theta)$ is equivalent to finding $g(\lambda,\theta)$ to make the expected value of $I(\lambda)/g(\lambda,\theta)$ be constant; this is asymptotically equivalent to finding a filter such that the filtered series is a white noise. Therefore, for ARMA processes, the estimate which minimizes $L'(\theta)$ is asymptotically equal to the least squares estimate discussed in Box and Jenkins (1970). However, the function $L'(\theta)$ is much easier to handle than most of the objective functions defined on the time domain, such as in Box and Jenkins (1970). The difficulty in estimating the parameters on the time domain is mainly due to the intractability of the maximum likelihood equations (even for Gaussian processes).

In the following sections, we study the properties of the weighted least squares estimate which minimizes the criterion function

$$Q_T(\theta) = \sum_\lambda \phi(\lambda)[f(\lambda,\theta)-I(\lambda)]^2. \tag{1.15}$$

We will show that, with a proper weighting function, the estimate has the same covariance matrix as that of the approximate maximum likelihood estimate.

The weighted least squares estimate has several advantages: (1) Due to the availability of various least squares algorithms, it is easy to implement the estimation procedure. (2) We are often only interested in the properties of the processes in certain frequency bands; by using a proper weighting function, we could avoid the effects from other bands or reduce the effect of additive noise. (3) The approximate maximum likelihood procedure cannot deal with the cases where $f(\lambda,\theta_0)$ vanishes at some λ. The weighted least squares procedure does not have this difficulty. (4). Analogous to the robust procedures for independent observations, we could get estimates resistant to the presence of peaks (corresponding to periodic components, such as seasonal effect) in the series. Therefore, it is possible to estimate the spectrum of the de-seasoned series without removing the trend from the series.

2. Assumptions and Results

We are interested in the series $X(t)$, $t=0,1, \ldots , T-1$, which satisfies the following assumption.

Assumption 1. $X(t)$ is a stationary series with cumulants

$$c_h(u_1,u_2, \ldots , u_{h-1}) = \text{cum}\{X(t+u_1), \ldots , X(t+u_{h-1}),X(t)\} \tag{2.1}$$

The cumulants satisfy

$$\sum_{u_1, \ldots , u_{h-1}=-\infty}^{\infty} (1+|u_j|)|c_h(u_1, \ldots , u_{h-1})| < \infty \tag{2.2}$$

for $j=1, \ldots , h-1$ and $h=2,3,\ldots.$

Under Assumption 1, $X(t)$ has h-th order spectrum $f_h(\lambda_1, \ldots , \lambda_{h-1})$ with bounded and uniformly

continuous derivatives (cf Brillinger (1975)).

Since the second-order properties of the series are often expressed in terms of its spectrum, it is more convenient to check whether $\sum[1+|u|]|c(u)| < \infty$ or not by inspecting the spectrum. A sufficient condition is given in the following theorem. The proofs of this and other theorems are given is Section 4.

Theorem 1. Suppose $X(t)$ is stationary and has spectrum $f(\lambda)$ with a bounded and uniformly continuous derivative $f'(\lambda)$ which satisfies a Lipschitz condition of order $\alpha > 0$. Then $\sum[1+|u|]|c(u)| < \infty$, here $c(u) = c_2(u)$ is the autocovariance function of the series.

All of the examples except (1.7) given in Section 1 satisfy the conditions of Theorem 1. The spectrum of example (1.7) is not differentiable at $\lambda = 0$, and the autocovariance function is of order u^{-2} (see Robinson (1978) and Whittle (1962)); thus Assumption 1 is not satisfied either. For situations similar to (1.7), we could filter the series to make the spectrum of the filtered series satisfy the conditions in Theorem 1. That is, we can use filter to smooth the singular points of the spectrum. Of course, we are not able to do this when the singular points depend on unknown parameters.

We now establish two theorems which are of some independent interest and are more general than the results required to prove the asymptotic properties of the weighted least squares estimates.

Theorem 2. Let $\psi(\lambda)$ be a continuous function on $[-\pi,\pi]$. If $X(t)$ satisfies Assumption 1, then

$$\lim \frac{1}{T}\sum_{\lambda}\psi(\lambda)I^k(\lambda) = \frac{k!}{2\pi}\int\psi(\lambda)f^k(\lambda)d\lambda \quad \text{almost surely,} \tag{2.3}$$

for $k \geq 1$. Here $f(\lambda)$ is the (second order cumulant) spectrum of $X(t)$.

Theorem 2 can be easily extended to the following result.

Corollary 1. Suppose $\psi(\lambda,\theta)$ is uniformly continuous on $(\lambda,\theta) \in [-\pi,\pi] \times \Theta$, and $X(t)$ satisfies Assumption 1. Then, for $k \geq 1$,

$$\lim \frac{1}{T}\sum_{\lambda}\psi(\lambda,\theta)I^k(\lambda) = \frac{k!}{2\pi}\int\psi(\lambda,\theta)f^k(\lambda)d\lambda \quad \text{almost surely,} \tag{2.4}$$

and the convergence is uniform over Θ.

The asymptotic distributions of the variables in Theorem 2 are given below.

Theorem 3. Under the conditions in Theorem 2,

$$\sqrt{T}\left[\frac{1}{T}\sum_{\lambda}\psi(\lambda)I^k(\lambda) - \frac{k!}{2\pi}\int_{-\pi}^{\pi}\psi(\lambda)f^k(\lambda)\mathrm{d}\lambda \right] \rightarrow N(0,\sigma_k^2) \tag{2.5}$$

where

$$\sigma_k^2 = (2\pi)^{-1}k^2(2k-2)!\int_{-\pi}^{\pi}\psi^2(\lambda)f^{2k}(\lambda)\mathrm{d}\lambda + (2\pi)^{-1}k^2(2k-2)!\int_{-\pi}^{\pi}\psi(\lambda)\psi(-\lambda)f^{2k}(\lambda)\mathrm{d}\lambda \tag{2.6}$$
$$+ (2\pi)^{-1}k^2(k!)^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\psi(\lambda_1)\psi(\lambda_2)f_4(\lambda_1,-\lambda_1,\lambda_2)f^{k-1}(\lambda_1)f^{k-1}(\lambda_2)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 \ .$$

Remark 1. For series with $f_4(\lambda_1,\lambda_2,\lambda_3)\equiv 0$, such as in the case of Gaussian processes, the third term of (2.6) vanishes, and the asymptotic distribution is the same as the one in which $I(\lambda_i)$ are true independent variables.

In proving the theorem concerning the strong convergence of the weighted least squares estimates, we need assumptions about Θ, $f(\lambda,\theta)$, and the weighting function $\phi(\lambda)$.

Assumption 2. $f(\lambda,\theta)$ is continuous on $(\lambda,\theta)\in[-\pi,\pi]\times\Theta$, and Θ is a compact set in R^r.

Assumption 3. $\phi(\lambda)$ is a symmetric continuous function on $[-\pi,\pi]$ and for all $\theta\neq\theta_0$, $\int_{-\pi}^{\pi}\phi(\lambda)[f(\lambda,\theta)-f(\lambda,\theta_0)]^2\mathrm{d}\lambda>0$.

Assumption 3 requires that, on the support of $\phi(\lambda)$, $f(\lambda,\theta_0)$ is different from $f(\lambda,\theta)$, $\theta\neq\theta_0$. From Corollary 1 and Theorem 2, the strong consistence of the weighted estimate can be established.

Theorem 4. Under Assumptions 1 to 3, the estimate $\hat{\theta}$ which minimizes $Q_T(\theta)$ of (1.15) converges to θ_0 almost surely.

We need some more assumptions to derive the asymptotic distribution of the weighted least squares estimate

Assumption 4. θ_0 is an interior point of Θ.

Assumption 5. In a neighborhood of θ_0, $f(\lambda,\theta)$ is twice differentiable with respect to θ, and the derivatives are continuous in λ and θ.

Assumption 6. The r by r matrix $A = A(\theta_0)$ is positive definite, and the ij-th element of $A(\theta)$ is

$$a_{ij}(\theta) = \frac{1}{2\pi} \int \phi(\lambda) f_i(\lambda,\theta) f_j(\lambda,\theta) d\lambda \tag{2.7}$$

where $f_i(\lambda,\theta)$ is the derivative of $f(\lambda,\theta)$ with respect to θ_i, $i=1,\ldots,r$.

We now describe the asymptotic distribution of the weighted least squares estimate.

Theorem 5. Under the conditions in Theorem 4, and Assumptions 4 to 6, let $\hat{\theta}$ be the estimate which minimizes $Q_T(\theta)$ of (1.15). Then $\sqrt{T}(\hat{\theta}-\theta_0)$ is asymptotically normal with mean zero and covariance matrix $2A^{-1}BA^{-1}+A^{-1}DA^{-1}$. Here A is given in Assumption 6. The ij-th element of B and D are, respectively,

$$b_{ij} = \frac{1}{2\pi} \int \phi^2(\lambda) f^2(\lambda,\theta_0) f_i(\lambda,\theta_0) f_j(\lambda,\theta_0) d\lambda \tag{2.8}$$

and

$$d_{ij} = \frac{1}{2\pi} \int \int \phi(\lambda)\phi(\mu) f_4(\lambda,-\lambda,\mu) f_i(\lambda,\theta_0) f_j(\lambda,\theta_0) d\lambda \quad . \tag{2.9}$$

Remark 2. If we let $\phi(\lambda) = f^{-2}(\lambda,\theta_0)$, we obtain the same covariance matrix as the one in Theorem 5 of Robinson (1978), which is the asymptotic covariance matrix of the estimate which maximizes the approximate likelihood function $L(\theta)$ of (1.12)

Remark 3. For Gaussian processes, $f_4(\lambda,\mu,\omega) \equiv 0$, the matrix D vanishes, and when $\phi(\lambda) = f^{-2}(\lambda,\theta_0)$, the covariance matrix is $2A^{-1}$, which is the asymptotic covariance matrix of the maximum likelihood estimate (see Hannon (1973)). Therefore, the estimate with weighting function $\phi(\lambda) = f^{-2}(\lambda,\theta_0)$ is asymptotically efficient.

Remark 4. An efficient estimate can be obtained by using an iterative procedure. We could use the unweighted least squares estimate to start with, or, use the weighted function $\phi(\lambda) = \hat{f}^{-2}(\lambda)$. Here $\hat{f}(\lambda)$ is an estimate of the spectrum obtained by smoothing the periodograms. We then let

$\phi(\lambda) = f^{-2}(\lambda, \hat{\theta}_0),$

where $\hat{\theta}_0$ is our initial estimate, and find the estimate $\hat{\theta}_1$. Usually, only two or three iterations are required.

The covariance matrix in Theorem 5 depends on the matrices A, B and D. A and B can be estimated by $A(\hat{\theta})$ and $B(\hat{\theta})$ respectively. An estimate for the matrix D is suggested by the following theorem.

Theorem 6. Let $\psi(\lambda)$ be a continuous function on $[-\pi, \pi]$, and let $X(t)$ satisfy Assumption 1, then

$$\frac{1}{T^2}\sum_\lambda\sum_\mu \psi(\lambda)\psi(\mu)I(\lambda)I(\mu) \tag{2.10}$$

converges, in meam square, to

$$(2\pi)^{-1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\psi(\lambda)\psi(\mu)f_4(\lambda,-\lambda,\mu)d\lambda d\mu + (2\pi)^{-1}\int_{-\pi}^{\pi}\psi^2(\lambda)f^2(\lambda)d\lambda \tag{2.11}$$
$$+ (2\pi)^{-1}\int_{-\pi}^{\pi}\psi(\lambda)\psi(-\lambda)f^2(\lambda)d\lambda + \left[(2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)f(\lambda)d\lambda\right]^2 .$$

3. Simulation

We conducted two simulation experiments to evaluate the performance of the estimate. We simulated autoregressive processes

$$X(t)-0.6X(t-1)+0.3X(t-2) = \epsilon(t) \tag{3.1}$$

where $\epsilon(t)$ are independent random variables with distribution $N(0,1)$. The size of the simulated series is 256. The sample size is 200 for both experiments.

In the first experiment, we use the weighted function $\phi(\lambda)=\hat{f}^{-2}(\lambda,\theta_0)$ to find the estimate $\hat{\theta}_T$. The second experiment uses an iterative procedure to find the estimates. The initial estimate $\hat{\theta}_0$ is the unweighted least squares estimate. The first iterative estimate $\hat{\theta}_1$ is the estimate using the weighting function $\phi(\lambda)=f^{-2}(\lambda,\hat{\theta}_0)$. The second iterative estimate $\hat{\theta}_2$ is the estimate using the weighting function $\phi(\lambda)=f^{-2}(\lambda,\hat{\theta}_1)$. Table 1 compares the sample means of the estimates $\hat{\theta}_T$, $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ with the true parameters. In Table 2, the sample covariance matrices are compared with the asymptotic covariance matrix of the estimate $\hat{\theta}_T$, which is also the asymptotic covariance

matrix of the maximum likelihood estimate. From Table 2, it can be seen that the sample covariance matrix of the second iterative estimate $\hat{\theta}_2$ is very close to the asymptotic covariance matrix of $\hat{\theta}_0$.

4. Proofs

We now give proofs of the results in Section 2.

Proof of Theorem 1. From integration by part, we have

$$iuc(u) = iu\int_{-\pi}^{\pi}f(\lambda)\exp(iu\lambda)d\lambda \tag{4.1}$$
$$= \left\{f(\pi)[\exp(iu\pi)-\exp(-iu\pi)]\right\} - \int_{-\pi}^{\pi}f'(\lambda)\exp(iu\lambda)d\lambda.$$

Since the first term in (4.1) is equal to zero, then

$$\sum_{u=-\infty}^{\infty}|uc(u)| = \sum_{u=-\infty}^{\infty}|\int_{-\pi}^{\pi}f'(\lambda)\exp(iu\lambda)d\lambda| < \infty \tag{4.2}$$

from Theorem (3.6) of Zygmund (1959) (p. 241). The theorem is finished by noting that $c(0)$ is bounded.

Proof of Theorem 2. Let

$$\psi_L(\lambda) = \sum_{u=-L}^{L}q(u)\exp(iu\lambda) \tag{4.3}$$

be the Cesaro sum of the Fourier series of $\psi(\lambda)$, and let $S_T = \sum_{\lambda}\psi_L(\lambda)I^k(\lambda)$. The expected value of $I^k(\lambda)$ is equal to

$$(2\pi T)^{-k}E\left[\prod_{i=1}^{2k}d_X^T(\lambda_i)\right] = (2\pi T)^{-k}\sum_{\nu}\text{cum}(d_X^T(\lambda_i):i\in\nu_1)\cdots\text{cum}(d_X^T(\lambda_i):i\in\nu_p) \tag{4.4}$$

where $\lambda_i = (-1)^i\lambda$, and the summation is over all partitions $\nu = \nu_1\bigcup\cdots\bigcup\nu_p$ of the set $\{1,2,\ldots,2k\}$. From Theorem 4.3.2 of Brillinger (1975),

$$\text{cum}\{d_X^T(\lambda_1),\ldots,d_X^T(\lambda_h)\} \tag{4.5}$$
$$= (2\pi)^{k-1}\Delta^{(T)}\left(\sum_{j=1}^{h}\lambda_j\right)f_h(\lambda_1,\ldots,\lambda_{h-1}) + O(1)$$

where

$$\Delta^{(T)}(\lambda) = \sum_{t=0}^{T-1} \exp(-i\lambda t) . \tag{4.6}$$

The function $\Delta^{(T)}(\lambda)$ has the properties: $\Delta^{(T)}(\lambda)=T$ for $\lambda\equiv 0 \pmod{2\pi}$, and $\Delta^{(T)}(2\pi s/T)=0$ for s an integer with $s\neq 0 \pmod{T}$. The product of the cumulants in (4.4) is of order T^k for the partitions consisting of k sets, each of which contains an even number and an odd number. The number of such partitions is $k!$, and the products of the cumulants are equal to

$$[\text{cum}(d_X^T(\lambda),d_X^T(-\lambda))]^k = (2\pi T)^k f^k(\lambda)+O(T^{k-1}) . \tag{4.7}$$

For other partitions, the orders of the products are smaller than T^k. So,

$$\lim \frac{1}{T}\text{E}[\sum_{\lambda}\psi_L(\lambda)I^k(\lambda)] = \frac{k!}{2\pi}\int_{-\pi}^{\pi}\psi_L(\lambda)f^k(\lambda)d\lambda. \tag{4.8}$$

In the proof of Theorem 3, we will show that the variance of $T^{-1}S_T$ is of order T^{-1}. This implies, by Chebyshev's inequality and the Borel-Cantelli Theorem, that $N^{-2}S_{N^2}$ converges almost surely to the right hand side of (4.8). To prove the theorem, it is sufficient to show that S_M does not differ much from the nearest S_{N^2}.

Let $B_T = (2\pi)^k T^{k-1}[S_T+\phi(0)I(0)]$, and note that

$$B_T = \sum_{u=-L}^{L} q(u)\sum_{\Omega_u^T}X(t_1)X(s_1) \cdots X(t_k)X(s_k) \tag{4.9}$$

where the second summation is over all $(t_1,s_1, \ldots , t_k,s_k)$ in

$$\Omega_u^T=\{(t_1,s_1, \ldots , t_k,s_k):\sum_{i=1}^{k}(t_i-s_i)\equiv u \bmod(T), 1\leq t_i,s_i\leq T\}. \tag{4.10}$$

Let

$$G_M=B_M-B_{N^2} = \sum_{u=-L}^{L} q(u) \sum_{\Omega_u^M-\Omega_u^{N^2}} X(t_1)X(s_1) \cdots X(t_k)X(s_k). \tag{4.11}$$

Then

$$\text{E}G_M^2 \leq \sum_{u=-L}^{L}\sum_{v=-L}^{L} q(u)q(v) \sum_{\Omega_u^M-\Omega_u^{N^2}}\sum_{\Omega_v^M-\Omega_v^{N^2}} \tag{4.12}$$

$$\text{E}\left\{ X(t_1)X(s_1) \cdots X(t_k)X(s_k)X(t'_1)X(s'_1) \cdots X(t'_k)X(s'_k)\right\}.$$

The expectation inside the summation of (4.12) is equal to

$$\sum_{\nu} \text{cum}(X(z_i); i \in \nu_1) \cdots \text{cum}(X(z_i); i \in \nu_p) \tag{4.13}$$

where the summation is over all partitions of $\{1, 2, \ldots, 4k\}$, and

$$z_i = \begin{cases} t_j, & j = (i+1)/2, \ i \text{ is odd and } i \leq 2k \\ s_j, & j = i/2, \ i \text{ is even and } i \leq 2k \\ t'_j, & j = (i+1)/2 - k, \ i \text{ is odd and } 2k < i \\ s'_j, & j = i/2 - k, \ i \text{ is even and } 2k < i \end{cases} \tag{4.14}$$

The partitions giving non-zero products must consist of sets with two or more items. The number of sets is therefore smaller or equal to $2k$. After rearranging the index in (4.12) according to the partitions, we see that $\text{E}G_M^2$ is of order N^{2k}. Therefore,

$$\text{E}[\max_{N^2 < M \leq (N+1)^2} G_M]^2 \leq \sum_{M=N^2+1}^{(N+1)^2} \text{E}G_M^2 = O(N^{2k+1}). \tag{4.15}$$

Hence $\text{E}\{\max_M [N^{-2}S_M - N^{-2}S_{N^2}]^2\}$ is of order N^{1-2k} for $k \geq 2$. For $k = 1$, it can be shown by straight calculation that the expectation is of order N^{-2}.

Finally, writing $T^{-1}S_T$ as

$$\frac{S_T}{T} = \frac{S_{N^2}}{N^2} + \left(\frac{S_T}{T} - \frac{S_T}{N^2} \right) + \left(\frac{S_T}{N^2} - \frac{S_{N^2}}{N^2} \right) \tag{4.16}$$

and applying Chebyshev's inequality and the Borel-Cantelli Theorem, we prove that

$$\lim T^{-1}S_T = \frac{k!}{2\pi} \int \psi_L(\lambda) f^k(\lambda) d\lambda \tag{4.17}$$

almost surely. The theorem now follows from the uniform convergence of $\psi_L(\lambda)$ to $\psi(\lambda)$ (cf Edward (1979), p. 87) and the almost sure convergence of $T^{-1}\sum_\lambda I^k(\lambda)$.

Proof of Corollary 1. Let

$$\psi_L(\lambda, \theta) = \sum_{u=-L}^{L} q(u, \theta) \exp(iu\lambda) \tag{4.18}$$

be the Cesaro sum of the Fourier series of $\psi(\lambda, \theta)$. Since

$$|q(u, \theta)| \leq |\int_{-\pi}^{\pi} \psi(\lambda, \theta) \exp(i\lambda u) d\lambda| \leq 2\pi \sup_{\lambda, \theta} |f(\lambda, \theta)|, \tag{4.19}$$

$q(u, \theta)$ is uniformly bounded. From this and the fact that $\psi_L(\lambda, \theta)$ is a linear combination of the

$2L+1$ functions

$$\phi(\lambda,u)=\left[1-\frac{|u|}{L+1}\right]\exp(i\lambda u), \quad u=-L,\ldots,L, \tag{4.20}$$

we establish the uniform convergence of

$$T^{-1}\sum_{\lambda}\psi_L(\lambda,\theta)I^k(\lambda). \tag{4.21}$$

The corollary follows from the uniform convergence of $\psi_L(\lambda,\theta)$ to $\psi(\lambda,\theta)$, which can be shown by slightly modifying the proof of 3.2.2 of Edwards (1979).

Proof of Theorem 3. The variance of $T^{-1/2}\sum_{\lambda}\psi(\lambda)I^k(\lambda)$ is equal to

$$T^{-1}\sum_{\lambda_1}\sum_{\lambda_2}\psi_L(\lambda_1)\psi_L(\lambda_2)\mathrm{cum}(I^k(\lambda_1),I^k(\lambda_2)) \tag{4.22}$$

$$= T^{-2k-1}(2\pi)^{-2k}\sum_{\nu}\sum_{\lambda_1}\sum_{\lambda_2}\psi_L(\lambda_1)\psi_L(\lambda_2)\mathrm{cum}\{d_X^T(\omega_{ij});ij\in\nu_1\}\cdots\mathrm{cum}\{d_X^T(\omega_{ij});ij\in\nu_p)\},$$

where $\omega_{ij}=(-1)^j\lambda_i$ and the summation of ν is over all indecomposable partitions of the table (cf Brillinger (1975))

$$\begin{matrix} (1,1)\cdots(1,2k) \\ (2,1)\cdots(2,2k) \end{matrix} \tag{4.23}$$

From (4.5), it can be seen that there are only three groups of indecomposable partitions giving non-zero limits:

(1). There is one set which contains two items from each row; one of the two items in the same row is from an odd column and the other is from an even column. Each of the other sets contains two items (one from an even column, the other from an odd column) from the same row. The number of this kind of partition is $k^2(k!)^2$, and the product of the cumulants is

$$(2\pi)^{2k+1}T^{2k-1}f_4(\lambda_1,-\lambda_1,\lambda_2)[f(\lambda_1)f(\lambda_2)]^{k-1}, \tag{4.24}$$

where $f_4(\lambda,\mu,\omega)$ is the fourth cumulant spectrum of $X(t)$.

(2). There are at least two sets which contain one item from each row; both items are from even columns. The cumulants of these two sets are of order T when the λ_i's satisfy the restraint

$\lambda_1 = -\lambda_2$. The other sets could contain two items from the same row (one from an even column and the other from an odd column) or from the different rows (must satisfy the same restraint). The number of such partitions is $k^2(2k-2)!$, and the product of the cumulants is $(2\pi T)^{2k} f^{2k}(\lambda_1)$.

(3) Same as (2), but the sets consisting of items from different rows contain one item from an even colum and one item from an odd colum. The cumulants of these two sets are of order T when the λ_i's satisfy the restraint $\lambda_1 = \lambda_2$. The number of such partitions is $k^2(2k-2)!$, and the product of the cumulants is $(2\pi T)^{2k} f^{2k}(\lambda_1)$.

Therefore The variance of $T^{-1/2} \sum_\lambda \psi(\lambda) I^k(\lambda)$ is equal to

$$2\pi k^2 (k!)^2 T^{-2} \sum_{\lambda_1} \sum_{\lambda_2} \psi(\lambda_1) \psi(\lambda_2) f(\lambda_1, -\lambda_1, \lambda_2) f^{k-1}(\lambda_1) f^{k-1}(\lambda_2) \tag{4.25}$$
$$+ k^2(2k-2)! T^{-1} \sum_\lambda \psi^2(\lambda) f^{2k}(\lambda) + k^2(2k-2)! T^{-1} \sum_\lambda \psi(\lambda) \psi(-\lambda) f^{2k}(\lambda)$$

which converges to the σ_k^2 of (2.6) in Theorem 3.

We next prove the asymptotic normality by showing that the cumulants of order bigger than 2 converge to zero. The hth-order cumulant is

$$\mathrm{cum}\{ T^{-1/2} \sum_{\lambda_1} \psi(\lambda) I^k(\lambda), \ldots, T^{-1/2} \sum_{\lambda_k} \psi(\lambda) I^k(\lambda) \tag{4.26}$$
$$= T^{(-hk-h/2)} (2\pi)^{-hk} \sum_\nu \sum_{\lambda_1} \sum_{\lambda_2} \psi_L(\lambda_1) \psi_L(\lambda_2) \mathrm{cum}\{ d_X^T(\omega_{ij}); ij \in \nu_1) \} \cdots \mathrm{cum}\{ d_X^T(\omega_{ij}); ij \in \nu_p) \},$$

where $\omega_{ij} = (-1)^j \lambda_i$ and the summation of ν is over all indecomposable partitions of the table

$$(1,1) \cdots (1,2k)$$
$$(2,1) \cdots (2,2k)$$
$$\cdots \tag{4.27}$$
$$(h,1) \cdots (h,2k)$$

For any indecomposable partition, we consider

$$T^{(-hk-h/2)} \sum_\nu \sum_{\lambda_1} \sum_{\lambda_2} \psi_L(\lambda_1) \psi_L(\lambda_2) \mathrm{cum}\{ d_X^T(\omega_{ij}); ij \in \nu_1) \} \cdots \mathrm{cum}\{ d_X^T(\omega_{ij}); ij \in \nu_p) \}. \tag{4.28}$$

For the partition ν, we can find a set ν_1 which contains items from different rows. If ν_1 does not contain items from every row, we can find another set ν_2 which contains some items from rows

different from the rows of the items in ν_1. If the union of ν_1 and ν_2 does not contain items from all rows, we can find a set ν_3 which contain some items from rows different from the rows of the items in ν_1 or ν_2. We repeat this until $U = \nu_1 \bigcup \cdots \bigcup \nu_n$ contains items from every row. Suppose, in these n sets, there are m sets whose corresponding $\Delta^{(T)}$ function does not introduce linear restraints. Each of these m sets must contain at least 4 items. The number of items in U is at least $h+n-1+2m$ and the number of sets of the partition is smaller or equal to $hk-(h+n-1+2m)/2+n$. Also, the number of linear restraints introduced by the corresponding $\Delta^{(T)}$ functions of the remaining ν_i is $n-m$. Therefore, (4.28) is of order T to the power

$$[hk-(h+n-1+2m)/2+n]+[h-n+m]-hk-h/2 = -(n-1)/2. \tag{4.29}$$

Some possibilities for the ν with $n=1$ are as follows:

(1) ν_1 contains at least two items from each row. Then, the number of sets is at most $(k-1)h+1$. Hence, (4.28) is of order T to the power of $(k-1)h+1+h-kh-h/2=1-h/2$.

(2) ν_1 contains just one item from some rows. Then, there must exist a set ν_2 which contains items from different rows. The restraint introduced by ν_2 is different from the restraint introduced by ν_1. So, (4.28) is of order T to the power $[kh-h/2+1]+[h-2]-kh-h/2=-1$.

(3) Same as (2), but the restraint introduced by ν_2 is the same as the restraint introduced by ν_1, the number of items in $\nu_1 \bigcup \nu_2$ is at least $2h$ and (4.28) is of order T to the power $[kh-h+2]+[h-1]-kh-h/2=1-h/2$.

Therefore, the hth-order cumulants of $T^{-1/2}\sum_{\lambda}\psi(\lambda)I^k(\lambda)$ (4.26) converge to zero for all $h>2$ and the proof is finished.

Proof of Theorem 4. From Theorem 2,

$$\lim \frac{1}{T}\sum_{\lambda}\phi(\lambda)I^2(\lambda) = 2(2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)f^2(\lambda,\theta_0)d\lambda \tag{4.30}$$

almost surely. From Corollary 1,

$$\lim \frac{1}{T}\sum_\lambda \phi(\lambda)f(\lambda,\theta)I(\lambda) = 2(2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)f(\lambda,\theta)f(\lambda,\theta_0)d\lambda \tag{4.31}$$

almost surely and uniformly on Θ. Therefore

$$\lim T^{-1}Q_T(\theta) = (2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)[f^2(\lambda,\theta)-2f(\lambda,\theta)f(\lambda,\theta_0)+2f^2(\lambda,\theta_0)]d\lambda \tag{4.32}$$
$$= (2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)[f(\lambda,\theta)-f(\lambda,\theta_0)]^2 d\lambda + (2\pi)^{-1}\int_{-\pi}^{\pi}\phi(\lambda)f^2(\lambda,\theta_0)d\lambda \ .$$
$$= Q(\theta)$$

Under Assumption 3, the minimum of $Q(\theta)$ is attained only at $\theta=\theta_0$. Let $\hat{\theta}_T$ be a sequence of estimates which minimizes $Q_T(\theta)$. Suppose that there exists a subsequence $\hat{\theta}_{T_n}$ converging to $\theta'\neq\theta_0$. Then

$$Q(\theta')=\lim T^{-1}Q_T(\hat{\theta}_{T_n})\leq\lim T^{-1}Q_T(\theta_0)=Q(\theta_0) \tag{4.32}$$

which is a contradiction, and the proof is finished.

Proof of Theorem 5. This follows immediately from Theorem 3 and Corollary 1 by using the classical argument in Jennrich (1969).

Proof of Theorem 6. The expected value of (2.10) is equal to

$$(2\pi T)^{-2}T^{-2}\sum_\lambda\sum_\mu\psi(\lambda)\psi(\mu)\mathrm{E}[d_X^T(\lambda)d_X^T(-\lambda)d_X^T(\mu)] \tag{4.34}$$

Since

$$\mathrm{E}[d_X^T(\lambda)d_X^T(-\lambda)d_X^T(\mu)d_X^T(-\mu)] = \mathrm{cum}\{d_X^T(\lambda),d_X^T(-\lambda),d_X^T(\mu),d_X^T(-\mu)\} \tag{4.35}$$
$$+ \mathrm{cum}\{d_X^T(\lambda),d_X^T(-\lambda)\}\mathrm{cum}\{d_X^T(\mu)d_X^T(-\mu)\}$$
$$+ \mathrm{cum}\{d_X^T(\lambda),d_X^T(\mu)\}\mathrm{cum}\{d_X^T(-\lambda),d_X^T(-mu)\}$$
$$+ \mathrm{cum}\{d_X^T(\lambda),d_X^T(-\mu)\}\mathrm{cum}\{d_X^T(-\lambda),d_X^T(\mu)\},$$

(4.34) is equal to

$$2\pi T^{-2}\sum_\lambda\sum_\mu\psi(\lambda)\psi(\mu)f_4(\lambda,-\lambda,\mu) + T^{-1}\sum_\lambda\psi^2(\lambda)f^2(\lambda) \tag{4.36}$$
$$+ T^{-1}\sum_\lambda\psi(\lambda)\psi(-\lambda)f^2(\lambda) + \left[T^{-1}\sum_\lambda\psi(\lambda)f(\lambda)\right]^2 + o(1),$$

which converges to (2.11) in Theorem 5. Using a similar argument in the proof of Theorem 3, we can show that the variance of (2.10) is of order T^{-1}, and finish the proof.

REFERENCES

[1] Aki, K. (1967). Scaling law of seismic spectrum. *J.Geophys. Res.*, **72**, 1217-1231.

[2] Aki, K. and Richards, P.G. (1980). *Quantitative Seismology Theory and Methods*, W.H. Freeman and Co., San Francisco.

[3] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, **60**, 217-226.

[4] Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

[5] Brillinger, D. R. (1975). *Time Series Data Analysis and Theory*, Holt, Rinehart, and Winston, New York.

[6] Davis, R.B. (1973). Asymptotic inference in stationary Gaussian series. *Adv. Appl. Prob.*, **5**, 469-497.

[7] Dzhaparidze, K.O. (1974). A new method for estimating spectral parameters of a stationary regular time series. *Theory Prob. Appl.*, **19**, 122-132.

[8] Edwards, R.E. (1979). *Fourier Series: A Modern Introduction*, Vol. 1, Second Edition, Springer-Verlag, New York.

[9] Hannon, E.J. (1973). The asymptotic theory of linear time series models. *J. Appl. Prob.*, **10**, 130-145.

[10] Ibragimov, I.A. (1967). On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory Prob. Appl.*, **12**, 115-119.

[11] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, **40**, 633-643.

[12] Lumley, J.L. and Panofsky, H.A. (1964). *The Structure of Atmospheric Turbulence*. Wiley-Interscience, New York.

[13] Pierson, W.J. and Moskowitz, L. (1964), A proposed spectral form for fully developed wind seas based on the similarity theory of S.A. Kitaigorodskii. *J. Geophys. Res.*, **69**, 5181-5190.

--   --

[14] Robinson, P.M. (1978). Alternative models for stationary stochastic processes. *Stoch. Proc. Appl.*, **8**, 141-152.

[15] Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**, 105-146.

[16] Taniguchi, M. (1981). An estimation procedure of parameters of a certain spectral density model. *J. R. Statist. Soc.*, B. **43**, 34-40.

[17] Whittle, P. (1961). Gaussian estimation in stationary time series. *Bull. Int. Statist. Inst.*, **39**, 105-130.

[18] Whittle, P. (1962). Topographic correlations, power-law covariance functions, and diffusion. *Biometrika*, **49**, 305-314.

[19] Zygmund, A. (1959). *Trigonometric Series*, Vol. 1, Cambridge University Press, Cambridge.

Department of Mathematical Sciences
Rice University
Houston, Texas 77251

	θ_0	$\hat{\theta}_T$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$
			Table 1		
σ^2	1	0.990	0.964	0.984	0.988
a_1	-0.6	-0.601	-0.614	-0.585	-0.599
a_2	0.3	0.304	0.313	0.287	0.301

Comparison of the sample means of the estimates to the values of the true parameter θ_0.

(×10^{-3})	Table 2				
(×10^{-3})	Theoretic	$\hat{\theta}_T$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$
Var($\hat{\sigma}^2$)	7.81	6.98	25.23	8.54	8.18
Var($\hat{a}_1$)	3.55	3.43	13.36	3.68	3.35
Var($\hat{a}_2$)	3.55	3.22	8.40	3.68	3.36
Cov($\hat{\sigma}^2,\hat{a}_1$)	0	-0.34	13.31	-0.12	0.54
Cov($\hat{\sigma}^2,\hat{a}_2$)	0	0.41	-9.33	0.54	-0.42
Cov($\hat{a}_1,\hat{a}_2$)	-1.64	-1.47	-8.03	-2.11	-1.51

Comparison of the sample covariance matrices of the estimates to the asymptotic covariance matrix of $\hat{\theta}_T$.